Indian J.Sci.Res. 17(2): 95 - 98, 2018 ISSN: 0976-2876 (Print)
ISSN: 2250-0138(Online)

APPROACHTOAUTOMATIONTESTFRAMEWORKFORBIGDATAPLATFORMUSINGOPENSOURCETECHNOLOGIES

[1]Dr.Vasanthi Kumari P, [2]Ashok GV
[1,2] Departmentof Computer Science and Engineering, Dayananda Sagar University, Bangalore

Abstract-
TheKeytoasuccessfulautomationeffortliesindesigningagoodandrobustautomationframework.Automationcanimprovequalityactivitiesandleadtohighertestingefficiency.Asweareadealingwithhugedataandexecutingonmultiplenodestherearehighchancesofhavingbaddataanddataqualityissuesateachstageofprocess.Datafunctionaltestingisperformedtoidentifythesedataissuesbecauseofcodingerrorsornodeconfigurationerrors.Testingshouldbeperformedateachofthethreephasesofbigdataprocessingtoensurethatdataisgettingprocessedwithoutanyerrors.Functionaltestingincludes,i)Loadingsourcedatafilesintohdfs,ii)Performmapreduceoperationsandiii)Extracttheoutputresultsfromhdfs.Issueoccursini)datamovingfromsourcesystemstoHadoopareincorrectdatacapturedfromsourcesystems,incorrectstorageofdataincompleteorincorrectreplication.ii)Issuesfaceddataprocessingorcodingissuesinmapreducejobs,jobsworkingcorrectlywhenruninstandalonenode,butworkingincorrectlywhenrunonmultiplenodes,incorrectaggregations,nodeconfigurations,andincorrectoutputformat.iii)issuesfacedwhilegeneratingreportsarereportdefinitionsnotsetasperthe requirement,reportdataissues,layoutandformattingissues.Automationpromisestofindmoredefectswithlesseffort.It'samustforeverynewproject.Themystiqueofautomationissostrongthatitevenseemsthatyounolongertowritetestcases,runmanualtesting,ordoimpactanalyses.Open-sourcetechnologiesarehelpingorganizationsacrossindustriesgainstrategicinsightsfromthetorrentsofdatathatnowflowthroughITsystems.Arobustbig-datavalidationframeworkcansignificantlyimprovehigh-volume,big-datatestinghelpingtofortifyqualityassurance,refineanalyticalthinkingandenhancetheoverallcustomerexperience.Thispapergivesabriefinsightofgeneraltestautomationframeworksanddescribestestautomationframeworkwhichhelpedinincreasingproductivitybysimplifyingtheprocessoftestscriptdevelopment.

Keywords:TestAutomationFramework;OpenSourcetools;Validations;Reporting;Visualization;Analysis

IIntroduction

Theever-increasingcomplexityandscaleofITapplicationshasmadequalityandreliabilityofparamountimportance.However,deliveringthesamethroughtestingisbecomingachallenge.Softwaretestautomationhasthecapabilitytoincreasetheoveralltestcoverage,whilealsoimprovingsoftwarequality.Intoday'senvironmentsoftwareindustryisfacedwithchallengesofdevelopingincreasinglycomplexapplicationsandproductsovershortertimeperiodsandatlowercosts,whilemaintainingconsistentlyhighquality.Thisneedforshortertimetomarketasksforsignificantlyreduceddevelopmentandtestingtime.Thismakestestautomationincreasinglynecessaryandevencritical.TestAutomationistheprocessofwritingprogramstodotestingthatwouldotherwisebedonemanually.Oncetestshavebeenautomatedtheycanberunrepeatedlyandquicklymakingitcosteffective,especiallyforproductswithlongmaintenancelife.Itconsistsofadaptable,object-orientedcomponentsthatensurequick-and-easyimplementationofcustomizedsolutions

Theprimaryobjectivesoftestautomationare:

- Reducetheoveralltestcycletimeandacceleratethetimetomarket.
- Reducecostofrepetitivemanualtesting.
- Increasethetestcoverageandreducerisk.
- Ensuringconsistentandthoroughtestingleadingtoimprovedqualityandreliability
- OptimizethetimeandcostbyrepeatingthetestsandHighertestcoveragelevels
- Maintainable&scalablescriptsandMinimizeleadtime

Thepaperpresentstheliteraturereviewinsection2,Problemstatementinsection3,Proposedworkinsection4,Resultsanddiscussionsinsection5followedbyConclusioninthefinalsection.

IILiteratureReview

V.Vaidhehi,ArunM.B[1],Theoutcomeofthisresearchisondataprocessingandstorageapproachesarefacingmanychallengesinmeetingthecontinuouslyincreasingdemandsofbigdata.MainfocusedonMapreduce,oneofthekeyapproachesformeetingthebigdatademandsthroughhighlyparallelprocessingonalargeamountofcommoditynodes.Challengesandsolutionsonfourdimensionslikedatastorage,analytics,onlineprocessingandprivacyandsecurity.

RabiulIslamJony,AhsanHabib,NabeelMohammed,RakibulIslamRony2[2],Thispaperidentifiespotential Bigdatausecase

[1]Corresponding Author

domainsfortelecomoperators.Italsopresentsthecorrespondingexamples,requireddatatypesandtypicalchallengestoimplementtheusecases.However,newerusecasesareemergingasoperatorsaremorefocusingontheirbigdatastrategy.Forexample,potentialusecasesbasedonlocationdataareemergingcurrently.Asaresult,locationbasedservicescanbeintroducedasaseparateusecasedomain.Asthefieldofbigdataisstillinagreatstateofchangeanddevelopment,thepossibilitiesforresearchinthisareaareextensiveandtheinterestonthefindingswillcertainlyriseifttheadoptionofbigdatarisesasexpected.

TheComputingBigDataReview2015[3]summarizestheresultsofacomprehensiveresearchprogramundertakenbyComputingduringthefirstquarterof2015.Thefindingsarecomparedwithtosepublishedin theComputingBigDataReview2014inordertoestablishthedegreetowhichthemarkethasdevelopedinthisshortspaceoftimeandthedegreetowhichBigDataandthedemocratizationofdataanalyticsischangingbusinessorganizations.Thereviewcontainssomeuniqueinsightsfromhigh-rankingITdecisionmakersintohoworganizationsareapproachinganddeployingBigDatasolutionsandsomeofthespecificchallengestheyface.Methodologytheresearchprojectwasconductedusingacombinationofqualitativeandquantitativemethods.

IIIProblemStatement

Asweareddealingwithhugedataandexecutingonmultiplenodestherearehighchancesofhavingbaddataanddataqualityissuesateachstageofprocess.Datafunctionaltestingisperformedtoidentifythesedataissuesbecauseofcodingerrorsornodeconfigurationerrors.Testingshouldbeperformedateachofthethreephasesofbigdataprocessingtoensurethatdataisgettingprocessedwithoutanyerrors.Functionaltestingincludes,

1.VerificationofPre-Hadoopprocessing

Datafromvarioussourceslikeweblogs,socialnetworksites,callogs,transactionaldataetcisextractedbasedonrequirementsandloadedintoHDFSbeforeprocessingitfurther.**Issues**:SomeoftheissueswhichwefaceduringthisfaceofthedatamovingfromsourcesystemstoHadoopareincorrectdatacapturedfromsourcesystems,incorrectstorageofdataincompleteorincorrectreplication.

2.VerificationofHadoopmapreduceprocessdataoutput

OncedataisloadedintoHDFSHadoopmapreduceprocessisruntoprocessthedatacomingfromdifferentprocess.

Issues:Someissuesthatwefaceduringthisphaseofthedataprocessingorcodingissuesinmapreducejobs,jobsworkingcorrectlywhenruninstandalonenode,butworkingincorrectlywhenrunonmultiplenodes,incorrectaggregations,nodeconfigurations,andincorrectoutputformat.

3.Verificationofdataextractandloadinto EWD.

OnceMap-Reduceprocessiscompletedanddataoutputfilesandgenerated,thisprocesseddataismovedtoenterprisedatawarehouseoranyothertransactionalsystemsdependingontherequirement.

Issues:Someissuesthatwefaceduringthisphaseincludeincorrectlyappliedtransformationrules,incorrectloadofHDFSfilesintoEDWandincompletedataextractfromHadoopHDFS.

4.VerificationofReports

Analyticalreportsaregeneratedusingreportingtoolsbyfetchingthedatafrom EDWorrunningqueriesonHive.

.**Issues**:Someofissuesfacedwhilegeneratingreportsarereportdefinitionsnotsetasperthe requirement,reportdataissues,layoutandformattingissues.Apartfromthisfunctionalverification non-functionaltestingincludingperformancetestingandfailovertestingneedtobeperformed.

FailoverTesting: Hadooparchitectureconsistsofanamenodeandhundredsofdatanodeshostedonseveralservermachinesandeachofthemareconnected.Therearechancesofnodefailureandsomeofthe HDFScomponentsbecomenon-functional.Someoffailurescanbenodefailure,datanodefailureandnetworkfailure.HDFSarchitectureisdesignedtodeletthesefailuresautomaticallyrecovertoprocessedwiththeprocessing.Failovertestingisanimportantfocusareainbigdataimplementationswiththeobjectivevalidatingtherecoveryprocessandtoensurethedataprocessinghappensseamlesslywhenswitched tootherdatanodes.SomevalidationsthatneedtobeperformeddduringfailovertestingarevalidatingthatcheckpointsofeditlogsandFsImageofnamenodearehappeningatdefinedintervals,recoveryofeditlogsandFsImagefilesofnamenode,nodatacorruptionbecauseofthenamenodefailure,datarecoverywhendatanodefailsandvalidatingthatreplicationisinitiatedwhenoneofdatanodefailsordatabecomecorrupted.RecoveryTimeObjective(RTO)andRecoveryPointObjective(RPO)metricsarecapturedduringfailovertesting.

IVProposedMethodology

Developingtherightarchitecturalframeworkfortestautomationwillresultinautomationcodethatcanbeusedforlongerperiodsoftimewithlessmaintenancethanasimplerecord/playbacksolution.Thistranslatestoasignificantsavingsoverthecourseoflongprojects,andtheprovidesabilitytomorethoroughlytestanapplicationwithlessmanpower.Theresultofthisarchitectureisreliableautomationcodewithscriptsthatcanlasttheentirelifeoftheproduct(notjusttheproject)andthatcanbeusedandenhancedbybusinessanalystswhohavelittletonoknowledgeofautomatedtesting.Thisapproachshouldbestartedattheplanningoftheautomationwork.Therequirementsshouldbecorrectlydefinedandfinalized.Theautomationrequirementscanbeasformalandspecificasthesoftwarerequirements,ortheycanbeasinformalasagreeingwhatmodulesorportionsoftheapplicationshouldandshouldnotbeautomatedandwhichpaths,particulartestcaseswilltakethroughtheapplication.

TestAutomationFrameworkisagenericframeworkbasedonhybridmodel.Itisareusableframeworkthatcanbeeasilyinvokedbyanautomatedtestsuiteforanapplicationundertest.Itoptimizestheeffortrequiredtoautomatethetestingoftheapplication.

Theframeworkhidestheunderlyingchallengeslikecommunicatingwiththeenvironment,database,configurationsetc.fromtheapplicationundertest.Theframeworkcanworkwithmultipleautomationtoolsandapplicationsundertest.

TheframeworkisdividedintofourlogicalsectionsandisacollectionofvariousResources,DriverScripts,FunctionlibrariesTestAutomationArchitectureandResultlibraries.Figure1givesproposedTestAutomationFramework.

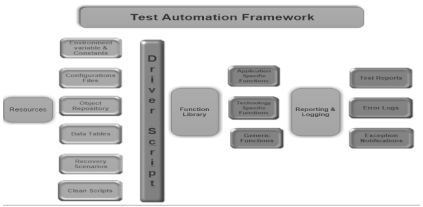

Figure1ProposedTestAutomationFramework

Resources: Thesearevariousinputsthatarerequiredforthedriverscriptstorun.Theymayconsistoftestdatafiles,objectrepositories,environmentvariablesetc.

FunctionLibraries: Theseformthecorepartoftheframeworkandconsistofreusablefunctions.Thesebasicallyrepresentbusinessflows,genericfunctionalitiesetc.andarecalledfromthedriverscripts.

DriverScripts: Theseconsistofcallstovariousreusablefunctionsandotherresourcesformingacompleteendtoendbusinessflow.

Reports: Theseconsistofresultsoftestexecution.

VResultsAndDiscussions

Bigdataisstillemergingandathereisalotofonusontesterstoidentifyinnovativeideastotesttheimplementation.Designedatestautomationframeworkin theBDT(BehaviorDrivenTesting)modelusingCucumber&GherkinsandJava.TheframeworkhelpedinperformingcountanddatavalidationduringthedataprocessingstagebycomparingtherecordcountbetweentheHiveandSQLtablesandconfirmedthatthedataisproperlyloadedwithoutanytruncationbyverifyingthedatabetweenHiveandSQLtables.

Similarly,whenitcomestovalidationonthemap-reduceprocessstage,ithelpsifthetesterhasgoodexperienceonprogramminglanguages.Thereasonisbecauseunlike SQLwhererequeriescanbeconstructedtoworkthroughthedataMapReduceframeworktransformsalistofkey-valuepairsintoalistofvalues.Agoodunittestingframeworklike Junitcanhelpvalidatetheindividualpartsof theMapReducejobbuttheydonottestthemasawhole.

Buildingatestautomationframeworkusingaprogramminglanguagelike Javacanhelphere.Theautomationframeworkcanfocusonthebiggerpicturepertainingto MapReducejobswhileencompassingtheunittestsaswell.Settinguptheautomationframeworktoacontinuousintegrationserverlike Jenkinscanbeevenmorehelpful.However,buildingtherightframeworkforbigdataapplicationsreliesonhowthetestenvironmentissetupasthepr ocessinghappensinadistributedmannerhere.Therecouldbeacl usterofmachinesontheQAserverwheretestingofMapReducejobsshouldhappen.

Theopensourcecommunityhasalsobeenhardatworkontoolsspecifictoanalyticsandbigdata.Afterthefoundationisbuiltusingopensourcetools,bigdatascientistshavetheoptionofcontinuingtouseopensourcetodrillintodataandproducereports.

DataTorrent,whichwascreatedbytheteambehindApacheApex,isonesuchtool.Itallowsforreal-timestreamingdataprocessingandbatchprocessing.It'sbuiltonHadoopandwasnamedasoneofthesevenleadersinbigdatastreaminganalyticssolutions.

Genieisanotheropensourcebigdatatool,thisonecreatedbyNetflixandalsorunningonHadoop.WhileNetflixcreatedGenieforitselforiginally.NetflixalsoreleasedLipstickasopensource,whichallowsforgraphicalrepresentationofHadoopPigjobs.

LinkedIndevelopedPinot,adistributedOLAPdatastorewithreal-timescalableanalytics,aswellasTaiga,whichreliesontheKafkamessagingsystemandHadoopYARNtoconductdistributedstreamprocessing.

Technology: Java1.8,Jenkins,SeleniumWebDriver,TestNG,Maven,Skool,DataTorrent,Genie,Lipstick,Pinot,Taiga,AutoIT,Winium,Sikuli.

VIConclusionAndFutureWork

Testautomationframeworkcouldbedifficulttoimplementinthebigdataplatform.Theframeworkunderlyingchallengeslikecommunicatingwiththeenvironment,database,configurationsetc,Proposedsolutionidentifyingtherequirementsandbuildingarobustautomationframeworkcanhelpindoingcomprehensivetesting.Inordertokeepupwiththepaceofproductdevelopmentanddeliveryitisessentialtoimplementeffective,reusabletestautomation.Properlyplannedandimplementedautomatedtestingcansignificantlylowerprojectriskandcost.Ifautomationiswelldesignedandimplemented,itwillnotonlyreducetheriskofthecurrentproject,butcanactuallyreducetheriskandcostoffutureprojects.TheROIfortestautomationiseasytoestablish.AutomationFrameworkprovidesawaytodriveproductivityandfostercodereuse,ultimatelyenhancingthequalityofresulting.However,alotwoulddependonhowtheskillsofthetesterandhowthebigdataenvironmentissetup.Andhaveidentifiedsomemajorchallengesregardingbigdatasupportingopensourcetools.Butthereare stillsomechallengeswithnomitigationstrategies.Ourfutureresearchwillconcentrateondevelopingamorecompleteunderstandingofchallengesassociatedwithbigdataandcomp

leteTestAutomationframeworkwhichincludesmaintenances ofperformancetesting.

References

[1] AuthorV.Vaidhehi,ArunM.B,DepartmentofComputerScience,ChristUniversityBengalurupublishedthepaperon“ManagingBigdatausingHadoopMapReduceinTelecomDomain”atInternationalJournalofInnovativeResearchinAdvancedEngineering(IJIRAE)ISSN:2349-2763Issue03,Volume3(March2016)www.ijirae.com

[2] AuthorsRabiulIslamJony,AhsanHabib,NabeelMohammed,RakibulIslamRony2publishedresearchpaperon‘BigDataUseCaseDomainsforTelecomOperators’fromPrimeasiaUniversity,Bangladeshin2015IEEEInternationalConference.

[3] TheComputingBigDataReview2015summarizestheresultsofacomprehensiveresearchprogrammerundertakenbyComputingduringthefirstquarterof2015.http://www.informationweek.com/big-data/big-data-analytics/big-data-predictions-for-2016/d/d-id/1323671https://www.splunk.com/en_us/solutions/solution-areas/big-data.htmlhttps://www.bigdatareviews.org/?p=977

[4] InMay2012,Elsevierssponsoreda2-dayconferenceinCanberra,Australiadedicatedtothetopicsof BigData,E-scienceandSciencepolicy(seevideosandlinkstothepresentationshere:https://www.youtube.com/playlist?list=PL61DD522B24108837.

[5] http://ieeexplore.ieee.org/

[6]. https://en.wikipedia.org/wiki/Big_data

[7]. InternationalJournalofInnovativeResearchinAdvancedEngineering(IJIRAE)ISSN:2349-2763Issue03,Volume3(March2016)

[8] shwarappa,AnuradhaJ,”ABriefIntroductiononBigData5V’sCharacteristicsandHadoopTechnology”,InternationalConferenceonIntelligentComputing,Communication&Convergence(ICCC-2015),

[9] BhartiThakur,ManishMann,”Dataminingforbigdata:AReview”,InternationaljournalofadvancedResearchinComputerScienceandSoftwareEngineering,ISSN:2277128x,Volume4,Issue5,May2014.

[10]. BigData:Survey,Technologies,Opportunities,and Challengeshttps://www.hindawi.com/journals/tswj/2014/712826/